

WHAT IS CLAIMED IS:

1. A method of preparing an artificial transcription factor (ATF) capable of modulating expression of a gene by interaction with a target site associated with said gene which
5 comprises

(a) preparing a combinatorial library of ATFs, each of said ATFs comprising a DNA-binding domain and a transcriptional regulatory domain, wherein said DNA-binding domain comprises three or more zinc fingers, wherein at least one of said zinc fingers has been rationally-designed so that the library contains at least one ATF for each of the 256 four-base-pair target sequences for one rationally-designed zinc finger;
10

(b) screening said library, a subset of members of said library or individual members of said library, or selecting for one or more members of said library, which modulate expression of said gene relative to a control level of expression;

(c) identifying gene expression modulating activity associated with the library, subset or member(s);
15

(d) optionally, subdividing the library or subset into smaller subsets or individual members and repeating steps (b) and (c); and

(e) recovering one or more ATFs having the desired gene expression modulating activity.
20

2. The method of Claim 1, wherein the transcriptional regulatory domain comprises a transcriptional activator or a protein domain which exhibits transcriptional activator activity.

3. The method of Claim 2, wherein said modulating activity is enhancing, increasing or up regulating transcription or gene expression.

4. The method of Claim 1 wherein the transcriptional regulatory domain comprises a transcriptional repressor or a protein domain which exhibits transcriptional repressor activity.
25

5. The method of Claim 4, wherein said modulating activity is repressing, reducing or down regulating transcription or gene expression.

6. The method of Claim 1 wherein the transcriptional regulatory domain comprises a transcription factor recruiting protein or a protein domain which exhibits transcription factor recruiting activity.
30

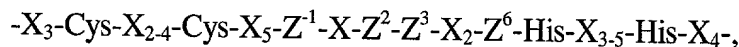
7. The method of Claim 6, wherein said modulating activity is enhancing, increasing or up regulating transcription or gene expression.

8. The method of Claim 6, wherein said modulating activity is repressing, reducing or down regulating transcription or gene expression.

5 9. The method of Claim 1 wherein said library contains 256^n members, wherein n is 1 to 6, and there are n rationally-designed zinc fingers in each ATF.

10. The method of Claim 1, wherein said target site for said ATF is unknown prior to said first screening or selecting step.

10 11. The method of Claim 1, wherein the DNA binding domain of said combinatorial library is prepared by a modular assembly method using at least one set of 256 oligonucleotides, each oligonucleotide comprising a nucleotide sequence encoding one of the 256 zinc fingers represented by the formula



wherein

15 X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, or glutamic acid;

Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

20 Z^6 is arginine, glutamine, threonine, or glutamic acid.

12. The method of Claim 11, wherein the X positions of said zinc finger domains comprise the corresponding amino acids from an Sp1, Sp1C or a Zif268 zinc finger.

13. The method of Claim 11, wherein said modular assembly method comprises

25 (a) preparing 256 individual mixtures or a single mixture of 256 members, under conditions for performing a polymerase-chain reaction (PCR), comprising:

(i) a first double-stranded oligonucleotide encoding a first zinc finger domain,

(ii) a second double-stranded oligonucleotide encoding a second zinc finger domain,

(iii) a third double-stranded oligonucleotide encoding a third zinc finger,

30 (iv) a first PCR primer complementary to the 5' end of the first oligonucleotide,

(v) a second PCR primer complementary to the 3' end of the third oligonucleotide,

wherein the 3' end of the first oligonucleotide is sufficiently complementary to the 5' end of the second oligonucleotide to prime synthesis of said second oligonucleotide therefrom,

wherein the 3' end of the second oligonucleotide is sufficiently complementary to the 5' end of the third oligonucleotide to prime synthesis of said third oligonucleotide therefrom,

wherein the 3' end of the first oligonucleotide is not complementary to the 5' end of the third oligonucleotide and the 3' end of the second oligonucleotide is not complementary to the 5' end of the first oligonucleotide, and

wherein when 256 individual mixtures are used

(i) said first double-stranded oligonucleotide in each mixture is a different member of the set of 256 separate oligonucleotides,

(ii) said second double-stranded oligonucleotide in each mixture is a different member of the set of 256 separate oligonucleotides, or

(iii) said third double-stranded oligonucleotide in each mixture is a different member of the set of 256 separate oligonucleotides; and

wherein when a single mixture is used

(1) one of said first, second or third sets of double-stranded oligonucleotides is said set of 256 separate oligonucleotides and the remaining sets of double-stranded oligonucleotides can be all the same or all different;

(b) subjecting the mixture or mixtures to a PCR; and

(c) recovering the nucleic acid encoding the three zinc finger domains, either separately or as a mixture, and preparing nucleic acid encoding said DNA-binding domain.

14. The method of Claim 13, wherein any two or all three sets of first, second or third sets of double-stranded oligonucleotides is a set of 256 separate oligonucleotides, each oligonucleotide comprising a nucleotide sequence encoding one of the 256 zinc fingers represented by the formula

$-X_3\text{-Cys-}X_{2,4}\text{-Cys-}X_5\text{-Z}^1\text{-X-Z}^2\text{-Z}^3\text{-X}_2\text{-Z}^6\text{-His-}X_{3,5}\text{-His-}X_4\text{-},$

wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, or glutamic acid;

Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, or glutamic acid.

15. The method of Claims 13 or 14, wherein said nucleic acid encoding said DNA-binding domain is operatively linked to a nucleic acid encoding said transcriptional regulatory domain.

16. The method of Claim 13 or 14, wherein the first and second PCR primers independently include a restriction endonuclease recognition site.

17. One or more host cells comprising an expression vector comprising a member of the combinatorial library of any one of Claims 1, 2, 4, 6, 9, 11, 13 or 14.

18. The host cells of Claim 17, wherein a sufficient number of host cells are present to statistically represent at least 50% of the members of said combinatorial library.

19. The host cells of Claim 18, wherein said sufficient number statistically represents at least 60%, 70%, 80%, 90% or 100% of the members of said combinatorial library.

20. A method of preparing an artificial transcription factor (ATF) capable of modulating expression of a gene by interaction with a target site associated with said gene which comprises

(a) preparing a scanning library of ATFs, each of said ATFs comprising a DNA-binding domain and a transcriptional regulatory domain,

wherein said DNA-binding domain comprises X zinc fingers, wherein each of the X zinc fingers has been rationally-designed to bind to $(3X+1)$ consecutive base pairs of a nucleic acid of length N base pairs, with there being one ATF for each $(3X+1)$ consecutive base pairs that occurs at an interval of Y bases in said nucleic acid,

wherein X is 3 to 6,

Y is 1 to 10, and

N is greater than or equal to 20

(b) screening said library, a subset of members of said library or individual members of said library, or selecting for one or more members of said library, which modulate expression of said gene relative to a control level of expression;

(c) identifying gene expression modulating activity associated with the library, subset or member(s);

(d) optionally, subdividing the library or subset into smaller subsets or individual members and repeating steps (b) and (c); and

(e) recovering one or more ATF having the desired gene expression modulating activity.

21. The method of Claim 20, wherein N is selected from the group consisting of 30, 50, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000 and 5000.

22. The method of Claim 20, wherein Y is 1 or 2.

23. The method of Claim 20, wherein X is 3.

24. The method of Claim 20, wherein the transcriptional regulatory domain comprises a transcriptional activator or a protein domain which exhibits transcriptional activator activity.

25. The method of Claim 24, wherein said modulating activity is enhancing, increasing or up regulating transcription or gene expression.

26. The method of Claim 20 wherein the transcriptional regulatory domain comprises a transcriptional repressor or a protein domain which exhibits transcriptional repressor activity.

27. The method of Claim 26, wherein said modulating activity is repressing, reducing or down regulating transcription or gene expression.

28. The method of Claim 20, wherein the transcriptional regulatory domain comprises a transcription factor recruiting protein or a protein domain which exhibits transcription factor recruiting activity.

29. The method of Claim 28, wherein said modulating activity is enhancing, increasing or up regulating transcription or gene expression.

30. The method of Claim 28, wherein said modulating activity is repressing, reducing or down regulating transcription or gene expression.

31. The method of Claim 20, wherein said target site for said ATF is unknown prior to said first screening or selecting step.

32. The method of Claim 20, wherein the DNA binding domain of said scanning library is prepared by a modular assembly method using at least one set of 256 oligonucleotides, each oligonucleotide comprising a nucleotide sequence encoding one of the 256 zinc fingers represented by the formula

5 $-X_3\text{-Cys-}X_{2-4}\text{-Cys-}X_5\text{-Z}^1\text{-X-Z}^2\text{-Z}^3\text{-X}_2\text{-Z}^6\text{-His-}X_{3-5}\text{-His-}X_4\text{-}$,

wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, or glutamic acid;

10 Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, or glutamic acid.

33. The method of Claim 32, wherein the X positions of said zinc finger domains comprise the corresponding amino acids from an Sp1, Sp1C or a Zif268 zinc finger.

15 34. One or more host cells comprising an expression vector comprising a member of the scanning library of any one of Claims 20, 21, 24, 26, 28 or 32.

35. The host cells of Claim 34, wherein a sufficient number of host cells are present to statistically represent at least 50% of the members of said scanning library.

20 36. The host cells of Claim 35, wherein said sufficient number statistically represents at least 60%, 70%, 80%, 90% or 100% of the members of said scanning library.

37. A method of preparing a protein having or controlling a predetermined biological activity and further capable of interacting with a target site on a DNA which comprises

(a) preparing a combinatorial library of proteins, each of said proteins comprising a DNA-binding domain, wherein said DNA-binding domain comprises three or more zinc
25 fingers, wherein at least one of said zinc fingers has been rationally-designed so that the library contains at least one protein for each of the 256 four-base-pair target sequences for one rationally-designed zinc finger;

(b) screening said library, a subset of members of said library or individual members of said library, or selecting for one or more members of said library, which exhibit or control said
30 predetermined biological activity relative to a control level of said biological activity;

(c) identifying said biological activity or control of said biological activity associated with the library, subset or member(s);

(d) optionally, subdividing the library or subset into smaller subsets or individual members and repeating steps (b) and (c); and

(e) recovering one or more proteins having or controlling said biological activity.

38. A method of preparing a protein having or controlling a predetermined biological activity and capable of interacting with a target site on a nucleic acid which comprises

(a) preparing a scanning library of said proteins, each of said proteins comprising a DNA-binding domain,

wherein said DNA-binding domain comprises X zinc fingers, wherein each of the X zinc fingers has been rationally-designed to bind to $(3X+1)$ consecutive base pairs of a nucleic acid of length N base pairs, with there being one protein for each $(3X+1)$ consecutive base pairs that occurs at an interval of Y bases in said nucleic acid,

wherein X is 3 to 6,

Y is 1 to 10, and

N is greater than or equal to 20

(b) screening said library, a subset of members of said library or individual members of said library, or selecting for one or more members of said library, which exhibit or control said predetermined biological activity relative to a control level of said biological activity;

(c) identifying said biological activity or control of said biological activity associated with the library, subset or member(s);

(d) optionally, subdividing the library or subset into smaller subsets or individual members and repeating steps (b) and (c); and

(e) recovering one or more proteins having or controlling said biological activity.

39. The method of Claim 38, wherein N is selected from the group consisting of 30, 50, 100, 200, 300, 400, 500, 1000, 2000, 3000, 4000 and 5000.

40. The method of Claim 37 or 38, wherein said protein comprises an effector domain.

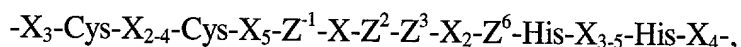
41. The method of Claim 40, wherein said effector domain comprises a transcriptional regulatory domain.

42. The method of Claim 37 or 38, wherein said effector domain comprises a transposase, integrase, recombinase, resolvase, invertase, protease, DNA methyltransferase, DNA demethylase, histone acetylase, histone deacetylase, nuclease, transcriptional repressor, transcriptional activator, single-stranded DNA binding protein, transcription factor recruiting protein, nuclear-localization signal, cellular uptake signal or any combination thereof.

43. The method of Claim 37 or 38, wherein said effector domain comprises a domain which exhibits transposase activity, integrase activity, recombinase activity, resolvase activity, invertase activity, protease activity, DNA methyltransferase activity, DNA demethylase activity, histone acetylase activity, histone deacetylase activity, nuclease activity, nuclear-localization signaling activity, transcriptional repressor activity, transcriptional activator activity, single-stranded DNA binding activity, transcription factor recruiting activity, cellular uptake signaling activity or any combination of such activities.

44. The method of Claim 37 or 38, wherein said target site for the DNA-binding domain is unknown prior to said first screening or selecting step.

45. The method of Claim 37 or 38, wherein the DNA binding domain of said combinatorial or scanning library is prepared by a modular assembly method using at least one set of 256 oligonucleotides, each oligonucleotide comprising a nucleotide sequence encoding one of the 256 zinc fingers represented by the formula



wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, or glutamic acid;

Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, or glutamic acid.

46. The method of Claim 45, wherein said modular assembly method comprises (a) preparing 256 individual mixtures or a single mixture of 256 members, under conditions for performing a polymerase-chain reaction (PCR), comprising:

(i) a first double-stranded oligonucleotide encoding a first zinc finger domain,

- (ii) a second double-stranded oligonucleotide encoding a second zinc finger domain,
(iii) a third double-stranded oligonucleotide encoding a third zinc finger,
(iv) a first PCR primer complementary to the 5' end of the first oligonucleotide,
(v) a second PCR primer complementary to the 3' end of the third oligonucleotide,

wherein the 3' end of the first oligonucleotide is sufficiently complementary to the 5' end of the second oligonucleotide to prime synthesis of said second oligonucleotide therefrom,

wherein the 3' end of the second oligonucleotide is sufficiently complementary to the 5' end of the third oligonucleotide to prime synthesis of said third oligonucleotide therefrom,

wherein the 3' end of the first oligonucleotide is not complementary to the 5' end of the third oligonucleotide and the 3' end of the second oligonucleotide is not complementary to the 5' end of the first oligonucleotide, and

wherein when 256 individual mixtures are used

- (i) said first double-stranded oligonucleotide in each mixture is a different member of the set of 256 separate oligonucleotides,
(ii) said second double-stranded oligonucleotide in each mixture is a different member of the set of 256 separate oligonucleotides, or
(iii) said third double-stranded oligonucleotide in each mixture is a different member of the set of 256 separate oligonucleotides; and

wherein when a single mixture is used

- (1) one of said first, second or third sets of double-stranded oligonucleotides is said set of 256 separate oligonucleotides and the remaining sets of double-stranded oligonucleotides can be all the same or all different;
(b) subjecting the mixture or mixtures to a PCR; and
(c) recovering the nucleic acid encoding the three zinc finger domains, either separately or as a mixture, and preparing nucleic acid encoding said DNA-binding domain.

47. The method of Claim 46 wherein said nucleic acid encoding said DNA-binding domain is operatively linked to a nucleic acid encoding said effector domain.

48. One or more host cells comprising an expression vector comprising a member of the combinatorial library of Claim 37 or 38.

49. The host cells of Claim 48, wherein said sufficient number statistically represents at least 50%, 60%, 70%, 80%, 90% or 100% of the members of said combinatorial library.

50. An isolated, artificial zinc finger protein (ZFP) for binding to a target nucleic acid sequence, said ZFP comprising at least three zinc finger domains, each zinc finger domain independently represented by the formula

$-X_3\text{-Cys-}X_{2-4}\text{-Cys-}X_5\text{-Z}^1\text{-X-Z}^2\text{-Z}^3\text{-}X_2\text{-Z}^6\text{-His-}X_{3-5}\text{-His-}X_4\text{-}$, said domains, independently, covalently joined to each other with from 0 to 10 amino acid residues;

wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, methionine or glutamic acid;

Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, tyrosine, leucine or glutamic acid;

provided that said protein does not have an amino acid sequence consisting of any one of SEQ ID. NOS. 3-12.

51. A nucleic acid comprising a nucleotide sequence encoding a ZFP of Claim 50.

52. An expression vector comprising the nucleic of Claim 51.

53. A host cell comprising the expression vector of Claim 52.

54. A method of preparing a zinc finger protein which comprises

(a) culturing the host cell of Claim 53 for a time and under conditions to express said ZFP; and

(b) recovering said ZFP.

55. An isolated fusion protein comprising

(a) a first segment which is a ZFP of Claim 50, and

(b) a second segment comprising a transposase, integrase, recombinase, resolvase, invertase, protease, DNA methyltransferase, DNA demethylase, histone acetylase, histone deacetylase, nuclease, transcriptional repressor, transcriptional activator, single-stranded DNA binding protein, transcription factor recruiting protein nuclear-localization signal or cellular uptake signal.

56. An isolated fusion protein comprising

(a) a first segment which is a ZFP of Claim 50, and

(b) a second segment comprising a protein domain capable of specifically binding to a binding moiety of a divalent ligand, said ligand capable of uptake by a cell.

57. An isolated fusion protein comprising

(a) a first domain encoding a single chain variable region of an antibody;

(b) a second domain encoding a nuclear-localization signal; and

(c) a third domain encoding transcriptional regulatory activity.

58. A nucleic acid comprising a nucleotide sequence encoding a ZFP of any one of Claims 55-57.

59. An expression vector comprising the nucleic of Claim 58.

60. A host cell comprising the expression vector of Claim 59.

61. A method of preparing a zinc finger protein which comprises

(a) culturing the host cell of Claim 60 for a time and under conditions to express said ZFP; and

(b) recovering said ZFP.

62. A method of making a nucleic acid encoding a zinc finger protein (ZFP) comprising three zinc fingers domains, each domain independently represented by the formula

-X₃-Cys-X₂₋₄-Cys-X₁₂-His-X₃₋₅-His-X₄-,

and said domains, independently, covalently joined with from 0 to 10 amino acid residues

which comprises:

(a) preparing a mixture, under conditions for performing a polymerase-chain reaction
5 (PCR), comprising:

(i) a first double-stranded oligonucleotide encoding a first zinc finger domain,

(ii) a second double-stranded oligonucleotide encoding a second zinc finger
domain,

(iii) a third double-stranded oligonucleotide encoding a third zinc finger,

10 (iv) a first PCR primer complementary to the 5' end of the first oligonucleotide,

(v) a second PCR primer complementary to the 3' end of the third
oligonucleotide,

wherein the 3' end of the first oligonucleotide is sufficiently complementary to the 5'
end of the second oligonucleotide to prime synthesis of said second oligonucleotide therefrom,

15 wherein the 3' end of the second oligonucleotide is sufficiently complementary to the 5'
end of the third oligonucleotide to prime synthesis of said third oligonucleotide therefrom, and

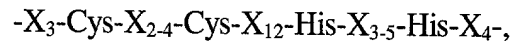
wherein the 3' end of the first oligonucleotide is not complementary to the 5' end of the
third oligonucleotide and the 3' end of the second oligonucleotide is not complementary to the
5' end of the first oligonucleotide;

20 (b) subjecting the mixture to a PCR; and

(c) recovering the nucleic acid encoding the three zinc finger domains and preparing a
nucleic acid encoding said ZFP.

63. The method of Claim 62, wherein the first and second PCR primers independently
include a restriction endonuclease recognition site.

64. A method of making a nucleic acid encoding a zinc finger protein (ZFP) having four or more zinc fingers domains, each domain independently represented by the formula



and said domains, independently, covalently joined with from 0 to 10 amino acid residues

which comprises:

(a) preparing a first nucleic acid according to the method of Claim 63, wherein said second PCR primer includes a first restriction endonuclease recognition site;

(b) preparing a second nucleic acid according to the method of Claim 63,

wherein said first and second PCR primers are complementary to the 5' and 3' ends, respectively, of the number of zinc finger domains selected for amplification,

wherein said first PCR primer includes a restriction endonuclease recognition site that, when subjected to cleavage by its corresponding restriction endonuclease, produces an end having a sequence which is complementary to and can anneal to, the end produced when said second PCR primer of step (a) is subjected to cleavage by its corresponding restriction endonuclease and

wherein said second PCR primer of step (b), optionally, includes a second restriction enzyme recognition site that, when subjected to cleavage produces an end that differs from and is not complementary to that produced from the first restriction endonuclease recognition site;

(c) optionally, preparing one or more additional nucleic acids by the method of Claim 63,

wherein said first and second PCR primers are complementary to the 5' and 3' ends, respectively, of the number of zinc finger domains selected for amplification,

wherein said first PCR primer for each additional nucleic acid includes a restriction endonuclease recognition site that, when subjected to cleavage by its corresponding restriction endonuclease, produces an end having a sequence which is complementary to and can anneal to the end produced when the second PCR primer used for preparation of the second nucleic

acid, or for the additional nucleic acid that is immediately upstream of the additional nucleic acid, is subjected to cleavage by its corresponding restriction endonuclease, and

wherein said second PCR primer for each additional nucleic acid, optionally, includes a restriction endonuclease recognition site that, when subjected to cleavage produces an end that differs from and is not complementary to any previously used;

(d) cleaving said first nucleic acid, said second nucleic acid and said additional nucleic acids, if prepared, with their corresponding restriction endonucleases to produce cleaved first, second and additional, if prepared, nucleic acids; and

(e) ligating said cleaved first, second and additional, if prepared, nucleic acids to produce the nucleic acid encoding a zinc finger protein (ZFP) having four or more zinc fingers domains.

65. An expression vector comprising a nucleic acid prepared by the method of any one of Claims 62-64.

66. A host cell comprising the expression vector of Claim 65.

67. A method of preparing a zinc finger protein which comprises

(a) culturing the host cell of Claim 66 for a time and under conditions to express said ZFP; and

(b) recovering said ZFP.

68. A method of designing a zinc finger domain of the formula

$-X_3\text{-Cys-X}_{2,4}\text{-Cys-X}_5\text{-Z}^1\text{-X-Z}^2\text{-Z}^3\text{-X}_2\text{-Z}^6\text{-His-X}_{3,5}\text{-His-X}_4\text{-}$,

wherein X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain which method comprises:

(a) identifying a target nucleic acid sequence having four bases;

(b) determining the identity of each X;

(c) determining the identity of amino acids at positions Z^1 , Z^2 , Z^3 and Z^6 as follows:

- (i) if the first base is G, then Z^6 is arginine or lysine,
 if the first base is A, then Z^6 is glutamine or asparagine,
 if the first base is T, then Z^6 is threonine, tyrosine, leucine, isoleucine or methionine,
 if the first base is C, then Z^6 is glutamic acid or aspartic acid,
- (ii) if the second base is G, then Z^3 is histidine or lysine,
 if the second base is A, then Z^3 is asparagine or glutamine,
 if the second base is T, then Z^3 is serine, alanine or valine,
 if the second base is C, then Z^3 is aspartic acid or glutamic acid,
- (iii) if the third base is G, then Z^1 is arginine or lysine,
 if the third base is A, then Z^1 is glutamine or asparagine,
 if the third base is T, then Z^1 is threonine, methionine leucine or isoleucine,
 if the third base is C, then Z^1 is glutamic acid or aspartic acid,
- (iv) if the complement of the fourth base is G, then Z^2 is serine or arginine,
 if the complement of the fourth base is A, then Z^2 is asparagine or glutamine,
 if the complement of the fourth base is T, then Z^2 is threonine, valine or alanine, and
 if the complement of the fourth base is C, then Z^2 is aspartic acid or glutamic acid; and

(d) preparing a zinc finger protein comprising said zinc finger domain.

69. A method of designing a multi-domained zinc finger protein (ZFP), each zinc finger domain independently represented by the formula

-X₃-Cys-X₂₋₄-Cys-X₅-Z¹-X-Z²-Z³-X₂-Z⁶-His-X₃₋₅-His-X₄-,

wherein X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain which method comprises:

(a) identifying a target nucleic acid sequence of length 3N+1 base pairs, wherein N is the number of overlapping 4 base pair segments of step (b);

(b) dividing said target nucleic acid sequence into overlapping 4 base pair segments, wherein the fourth base of each segment, up to the N-1 segment, is the first base of the immediately following segment;

(c) designing a zinc finger domain for each 4 base pair segment by

(i) determining the identity of each X; and

(ii) determining the identity of amino acids at positions Z¹, Z², Z³ and Z⁶ as follows:

(1) if the first base is G, then Z⁶ is arginine or lysine,

if the first base is A, then Z⁶ is glutamine or asparagine,

if the first base is T, then Z⁶ is threonine, tyrosine, leucine, isoleucine or methionine,

if the first base is C, then Z⁶ is glutamic acid or aspartic acid,

(2) if the second base is G, then Z³ is histidine or lysine,

if the second base is A, then Z³ is asparagine or glutamine,

if the second base is T, then Z³ is serine alanine or valine,

if the second base is C, then Z³ is aspartic acid or glutamic acid,

(3) if the third base is G, then Z¹ is arginine or lysine,

if the third base is A, then Z¹ is glutamine or aspartic acid,

if the third base is T, then Z¹ is threonine, methionine, leucine or isoleucine,

if the third base is C, then Z^1 is glutamic acid or aspartic acid,

- (4) if the complement of the fourth base is G, then Z^2 is serine or arginine,

if the complement of the fourth base is A, then Z^2 is asparagine or glutamine,

if the complement of the fourth base is T, then Z^2 is threonine, valine or alanine, and

if the complement of the fourth base is C, then Z^2 is aspartic acid or glutamic acid; and

- (d) preparing a ZFP comprising N zinc finger domains.

70. A method of binding a target nucleic acid with an artificial zinc finger protein (ZFP) which comprises contacting a target nucleic acid with a ZFP of Claim 50 in an amount and for a time sufficient for said ZFP to bind to said target nucleic acid.

71. A method of binding a target nucleic acid with a multi-domained zinc finger protein (ZFP) which comprises contacting a target nucleic acid of length $3N+1$ base pairs, wherein N is the number of overlapping 4 base pair segments in said target nucleic acid and wherein the fourth base of each segment, up to the N-1 segment, is the first base of the immediately following segment, with an amount of a multi-domained ZFP prepared according to the method of Claim 69 and for a time sufficient for said ZFP to bind to said target nucleic acid.

72. A method of modulating expression of a gene which comprises contacting a regulatory control element of said gene with a ZFP of Claim 50 in an amount and for a time sufficient for said ZFP to alter expression of said gene.

73. A method of modulating expression of a gene which comprises contacting a target nucleic acid in sufficient proximity to said gene with a fusion protein of a ZFP of Claim 50 fused to a transcriptional regulatory domain, wherein said fusion protein contacts said nucleic

acid in an amount and for a time sufficient for said transcriptional regulatory domain to alter expression of said gene.

74. A method of altering genomic structure which comprises contacting a target genomic site with a fusion protein of a ZFP of Claim 50 fused to a protein domain which exhibits transposase activity, integrase activity, recombinase activity, resolvase activity, invertase activity, protease activity, DNA methyltransferase activity, DNA demethylase activity, histone acetylase activity, histone deacetylase activity or nuclease activity, wherein said fusion protein contacts said target genomic site in an amount and for a time sufficient to alter genomic structure in or near said site.

75. A method of inhibiting viral replication which comprises

(a) introducing into a cell a nucleic acid encoding a ZFP of Claim 50, wherein said ZFP is competent to bind to a target site required for viral replication, and

(b) obtaining sufficient expression of said ZFP in said cell to inhibit viral replication.

76. A method of inhibiting viral replication which comprises

(a) introducing into a cell a nucleic acid encoding a fusion protein of a ZFP of Claim 50 fused to a single-stranded DNA binding protein, wherein said fusion protein is competent to bind to a target site required for viral replication, and

(b) obtaining sufficient expression of said fusion protein in said cell to inhibit viral replication.

77. A method of modulating expression of a gene which comprises

(a) contacting a eukaryotic cell with a divalent ligand capable of entry into said cell and comprising a first and second switch moiety of different specificity, wherein said cell contains

(i) a first nucleic acid expressing a first fusion protein of a ZFP of Claim 50 fused to a protein domain capable of specifically binding said first switch moiety, wherein said ZFP is specific for a target site in proximity to said gene, and

(ii) a second nucleic acid expressing a second fusion protein comprising a first domain capable of specifically binding said second switch moiety, a second domain which is a nuclear localization signal and a third domain which is a transcriptional regulatory domain;

5 (b) allowing said cell sufficient time to form a complex comprising said divalent ligand, said first fusion protein and said second fusion protein, to translocate said complex into the nucleus of said cell, to bind to said target site and to thereby to alter expression of said gene.

78. An artificial transposase comprising a catalytic domain, a peptide dimerization domain and a ZFP domain wherein said ZFP domain is a ZFP of Claim 50.

10 79. The transposase of Claim 78, which additionally comprises a terminal inverted repeat binding domain.

80. A method of target-specific introduction of an exogenous gene into the genome of an organism which comprises:

15 (a) introducing into a cell a first nucleic acid encoding a transposase of Claim 79, wherein said ZFP domain binds a first genomic target; a second nucleic acid encoding a transposase of Claim 79, wherein said ZFP domain binds a second genomic target; and a third nucleic acid encoding said exogenous gene, wherein said exogenous gene is flanked by sequences capable of being bound by the terminal inverted repeat binding domain of said transposases; and

20 (b) forming a complex among the genome, the third nucleic acid, and the two transposases sufficient for recombination to occur and thereby introduce said exogenous gene into the genome of the organism.

81. A method of target-specific excision of an endogenous gene from the genome of an organism which comprises:

25 (a) introducing into a cell a first nucleic acid encoding a transposase of Claim 78, wherein said ZFP domain binds a first genomic target; a second nucleic acid encoding a

transposase of Claim 78, wherein said ZFP domain binds a second genomic target; and wherein the endogenous gene is flanked by said first and second genomic targets; and

(b) forming a complex among the genome and the two transposases sufficient for recombination to occur and thereby excise said endogenous gene from the genome of the organism.

82. A method for detecting an altered zinc finger recognition sequence which comprises:

(a) contacting a nucleic acid containing the zinc finger recognition sequence of interest with a ZFP of Claim 50 specific for said sequence and conjugated to a signaling moiety, said ZFP present in an amount sufficient to allow binding of said ZFP to said zinc finger recognition sequence if said sequence was unaltered; and

(b) detecting binding of said ZFP to the zinc finger recognition sequence and thereby to ascertain that said zinc finger recognition sequence is altered if said binding is diminished or abolished relative to binding of said ZFP to the unaltered sequence.

83. A method of diagnosing a disease associated with abnormal genomic structure which comprises

(a) isolating cells, blood or a tissue sample from a subject;

(b) contacting nucleic acid from said cells, blood or said sample with a protein comprising a ZFP of Claim 50, a signaling moiety and, optionally, a cellular uptake domain wherein said ZFP binds to a target site associated with said disease; and

(c) detecting the binding of said protein to said nucleic acid to thereby make the diagnosis.

84. A set of 256 separate oligonucleotides, each oligonucleotide comprising a nucleotide sequence encoding one of the 256 zinc finger domains represented by the formula

$-X_3\text{-Cys-X}_{2-4}\text{-Cys-X}_5\text{-Z}^1\text{-X-Z}^2\text{-Z}^3\text{-X}_2\text{-Z}^6\text{-His-X}_{3-5}\text{-His-X}_4\text{-}$,

wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, or glutamic acid;

Z^2 is serine, asparagine, threonine or aspartic acid;

5 Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, or glutamic acid.

85. A set of oligonucleotides for producing a nucleic acid encoding zinc finger proteins having three or more zinc finger domains, said set comprising three subsets of 256 separate oligonucleotides, each oligonucleotide comprising a nucleotide sequence encoding one of the 256 zinc finger domains represented by the formula

$-X_3\text{-Cys-}X_{2-4}\text{-Cys-}X_5\text{-}Z^1\text{-X-}Z^2\text{-}Z^3\text{-}X_2\text{-}Z^6\text{-His-}X_{3-5}\text{-His-}X_4\text{-}$,

wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

15 Z^1 is arginine, glutamine, threonine, or glutamic acid;

Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, or glutamic acid; and

wherein

20 the 3' end of the first subset oligonucleotides are sufficiently complementary to the 5' end of the second subset oligonucleotides to prime synthesis of said second subset oligonucleotides therefrom,

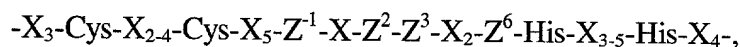
the 3' end of the second subset oligonucleotides are sufficiently complementary to the 5' end of the third subset oligonucleotides to prime synthesis of said third subset

25 oligonucleotides therefrom,

the 3' end of the first subset oligonucleotides are not complementary to the 5' end of the third subset oligonucleotides, and

the 3' end of the second subset oligonucleotides are not complementary to the 5' end of the first subset oligonucleotides.

- 5 86. A single-stranded or double-stranded oligonucleotide encoding a zinc finger domain for an artificial zinc finger protein (ZFP), wherein said oligonucleotide is from about 84 nucleotides to about 130 nucleotides and comprising a sequence encoding a zinc finger domain independently represented by the formula



- 10 and, optionally, a linker of from 0 to 10 amino acid residues;

wherein

X is, independently, any amino acid and X_n represents the number of occurrences of X in the polypeptide chain;

Z^1 is arginine, glutamine, threonine, methionine or glutamic acid;

- 15 Z^2 is serine, asparagine, threonine or aspartic acid;

Z^3 is histidine, asparagine, serine or aspartic acid; and

Z^6 is arginine, glutamine, threonine, tyrosine, leucine or glutamic acid.